Unpicking the Cause of Stereoselectivity in Actinorhodin Ketoreductase Variants with Atomistic Simulations

Stefano A. Serapian[†] and Marc W. van der Kamp*[,†,‡]

[†]School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, United Kingdom

[‡]Centre for Computational Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, United Kingdom

ABSTRACT: Ketoreductase enzymes (KRs) with a high degree of regio- and stereoselectivity are useful biocatalysts for the production of small, specific chiral alcohols from achiral ketones. Actinorhodin KR (*act*KR), part of a type II polyketide synthase involved in the biosynthesis of the antibiotic actinorhodin, can also turn over small ketones. *In vitro* studies assessing stereocontrol in *act*KR have found that, in the "reverse" direction, the wild-type (WT) enzyme's mild preference for *S*-*α*-tetralol is enhanced by certain mutations (*e.g.* P94L); and entirely reversed by others (*e.g.* V151L) in favor of *R*-*α*-tetralol. Here, we employ computationally cost-effective atomistic simulations to rationalize these trends in WT, P94L, and V151L *act*KR, using *trans*-1-decalone (**1**) as the model substrate. Three potential factors (**FI**-**FIII**) are investigated: frequency of *pro-R vs. pro-S* reactive poses (**FI**) is assessed with classical molecular dynamics (MD); binding affinity of *pro-R vs. pro-S* orientations (**FII**) is compared using the binding free energy method MM/PBSA; and differences in reaction barriers towards *trans*-1-decalol (**FIII**) are assessed by hybrid semiempirical quantum / classical (QM/MM) MD simulations with umbrella sampling, benchmarked with density functional theory. No single factor is found to dominate stereocontrol: **FI** largely determines the selectivity of V151L *act*KR, whereas **FIII** is more dominant in the case of P94L. It is also found that formation of *S*-*trans*-1-decalol or *R*-*trans*-1-decalol mainly arises from the reduction of the *trans*-1-decalone enantiomers *(4aS,8aR)*-**1** or *(4aR,8aS)*-**1**, respectively. Our work highlights the complexity of enzyme stereoselectivity as well as the usefulness of atomistic simulations to aid the design of stereoselective biocatalysts.

KEYWORDS: *Biocatalysis, Computational enzymology, QM/MM, Molecular dynamics, Ketoreductases, Chiral alcohols*

1. Introduction

Typically found in actinomycete bacteria, type II polyketide synthases (PKSs) are large multienzyme complexes consisting of multiple standalone units.[1-3] In a perfect example of enzymatic teamwork, each of these PKS domains is highly specialized in catalyzing a particular biosynthetic step, first "growing" a reactive polyketide chain, and then processing it into a variety of natural products: these include anticancer agents (*e.g.* daunorubicin and doxorubicin), and antibiotics (*e.g.* actinorhodin, tetracycline, and doxycycline).[1-2]

A key component of PKSs is their ketoreductase (KR) domain,[4-7] which typically exhibits some degree of regio- and stereoselectivity in catalyzing the reduction of polyketide chains. Together with several other KRs unrelated to PKSs, these enzymes have emerged as promising commercial biocatalysts[8-10] for the manufacture of small chiral alcohols from achiral ketones; and even for the synthesis of "unnatural natural products" with potentially improved antibiotic properties.[2] Such rise to prominence was significantly aided by the advent of novel protein reengineering techniques,[11-12] which have revolutionized biocatalysis by speeding up the design and screening of evolved KR mutants.[9-10]

Actinorhodin KR (*act*KR) from *Streptomyces coelicolor* has been attracting interest as a potential biocatalyst since at least the mid-2000s.[5, 13-14] Examination of its sequence and crystal structures[5, 13-14] confirmed the enzyme's homotetrameric nature (Figure 1a), and its structural and mechanistic similarity with enzymes of the short-chain dehydrogenase/reductase (SDR) family, especially fatty acid synthases (FASs). Essential for its activity are the presence in each active site of the NADPH cofactor, and of the catalytic tetrad Asn114-Ser144-Tyr157-Lys161 (Figure 1b).[13]

Another salient structural element is the flexible α6-α7 loop (Figure 1c and d), which varies substantially in size and composition across SDRs, and was proposed to play some role in the stereocontrol and activity of *Lactobacillus kefir* short-chain alcohol dehydrogenase.[10] Conformational variation of *act*KR's α6-α7 loop between "open" and "closed" is evident when superimposing monomers A and B of the wild-type *act*KR crystal structure (PDB: 2RH4; Figure 1d).[14]

The catalytic cycle of *act*KR begins when a ketone or polyketide substrate binds to its active site so that the oxygen of the carbonyl to be reduced is sandwiched between the catalytic Ser144 and Tyr157, with whose hydroxyl groups it forms hydrogen bonds (Figure 1b). Reduction then kicks off (Scheme 1; Figure 1b) when a hydride (H^-) from NADPH transfers to the carbonyl carbon, yielding the oxidized form $NADP^+$; concertedly (but asynchronously) with respect to the hydride transfer, the carbonyl oxygen abstracts a proton from Tyr157 (to which it remains hydrogen-bonded). The deprotonated Tyr157 is further stabilized by retaining its hydrogen bond to one of the cofactor's ribose hydroxyl groups (and is reprotonated in subsequent steps *via* H^+ transfers from the ribose and Lys161).[15] During this fundamental mechanistic step, which was shown to be rate-limiting in a human ketoacyl reductase,[15] the substrate's orientation with respect to the plane of NADPH's nicotinamide moiety determines whether the H^- attacks in a *pro-*<u>*R*</u> or *pro-*<u>*S*</u> fashion (Scheme 1 *top vs. bottom*), and thus ultimately dictates the stereochemical outcome of the reaction (<u>*R*</u>- or <u>*S*</u>- alcohol).

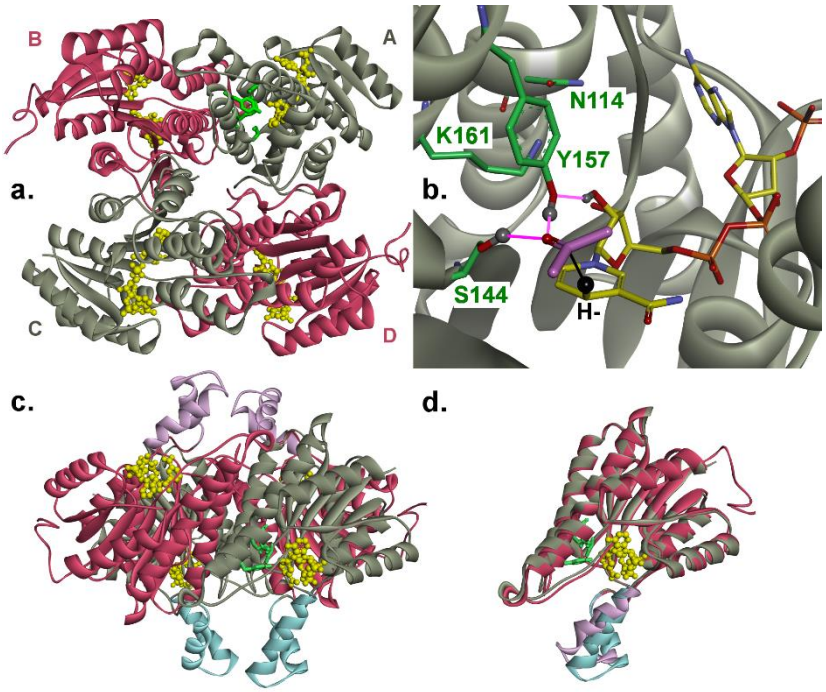
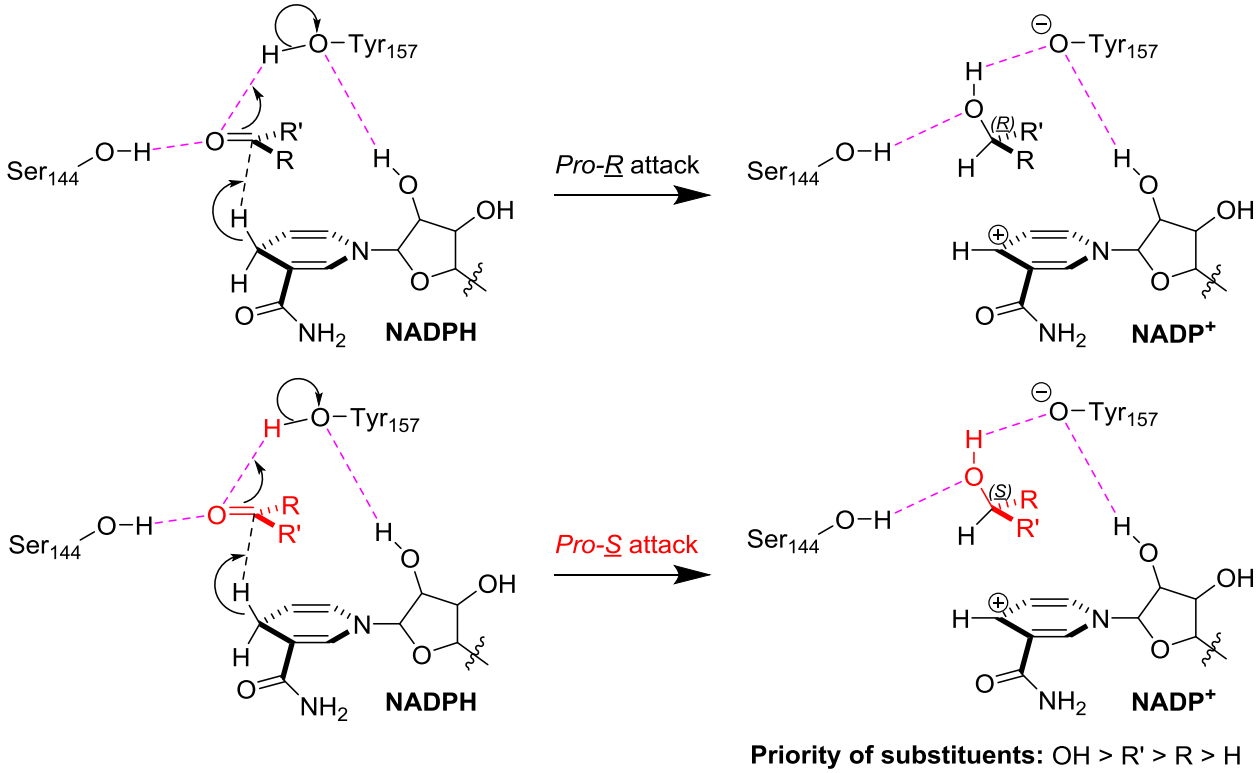


Figure 1. Structural features of wild-type actinorhodin ketoreductase (based on PDB ID 2RH4).[14] (*a*) Overall assembly of the four monomers A-D (grey: A, C; red: B, D) showing: the catalytic tetrad N114-S144-Y157-K161 in A (green sticks); and the NADPH cofactor in A-D (yellow spheres and sticks). H atoms omitted for clarity. (*b*) Closeup of the active site of monomer A, showing NADPH (C atoms in yellow); labelled tetrad residues (C in green); and a ketone fragment (C in magenta) docked in reactive position. O, N, P are rendered in red, blue, and orange, respectively. H atoms omitted for clarity, except for: hydroxyl H atoms on S144, Y157, NADPH ribose (grey spheres); and the reductive H^- (black). Relevant hydrogen bonds are shown in magenta; black line denotes direction of reductive H^- transfer. (*c*) Side view of *act*KR as shown in (*a*), highlighting the α6-α7 loop in monomers A-D (A, C, open: light blue; B, D, closed: pink). (*d*) Monomer B superimposed onto monomer A.

Scheme 1. Rate-limiting step of *act*KR (reductive hydride transfer) and its stereochemistry.[a,b]



[a] Ketone substrate orientation (left) determines whether H^- attack is *pro*-*R* or *pro*-*S*, and accordingly gives rise to an *R*- or *S*- product (right). [b] Magenta lines denote hydrogen bonds.

In nature, the enzyme typically forms a complex with an acyl carrier protein (ACP) bearing a 16-carbon polyketide chain (octaketide; *cf.* Chart 1 *top*), which is then unsheathed into one of the active sites, likely cyclized between the C7 and C12 positions and reduced specifically at the C9 position[14, 16-17] before further processing by other PKS modules.

In vitro, wild-type (WT) *act*KR and 29 strategically chosen mutants were extensively examined by Korman, Javidpour, *et al*.,[14, 16-17] looking for improved stereocontrol in the turnover of model ketone and alcohol substrates (*i.e.* in the "forwards" and "reverse" directions, respectively). More specifically, to study the "forwards" reduction reaction, the authors' preferred substrate is *trans*-1-decalone (**1**; Chart 1, *left*),[14, 16-17] since other potential candidates, notably the part-aromatic α-tetralone, were found to be turned over very sluggishly.[14] Rather than employing **1**'s directly corresponding alcohol *trans*-1-decalol (**2**) with its four possible stereoisomers (Chart 1, *center*), the authors assess the "reverse" oxidation reaction using the diastereomeric pair *R*-*α*-tetralol / *S*-*α*-tetralol (*R*-**3** / *S*-**3**; Chart 1, *right*),[14, 16-17] both of which are readily available commercially (in contrast to stereoisomers of **2**).

Assuming that the selectivity of the "forwards" and "reverse" reactions are still directly comparable microscopically (despite the slight change in substrates), two of the 29 mutants demonstrate particularly high stereoselectivity (as *per* data in Table 1): P94L has exclusive specificity for *S*-**3**; and V151L, has exclusive specificity for *R*-**3**. By comparison, WT *act*KR only has a very mild preference for *S*-**3** over *R*-**3** (3.5:1).

Table 1. *In vitro* specificity constants of key *act*KR variants towards reduction of 1 and oxidation of 3.

Variant	k_{cat}/K_M (**1**)[a,b]	k_{cat}/K_M (*R*-**3**)[a]	k_{cat}/K_M (*S*-**3**)[a]
WT[c,d]	3.23 ± 0.32	0.010 ± 0.001	0.035± 0.006
V151L[e]	1.28 ± 0.18	0.026 ± 0.005	inhibition
P94L[d]	1.02 ± 0.59	inhibition	0.036 ± 0.011

[a] s^{-1} mM^{-1}. [b] Racemate: *(4aS,8aR)*-**1** and *(4aR,8aS)*-**1** (*cf.* text). [c] Reference [14]. [d] Reference [16]. [e] Reference [17].

The "forwards" reaction is further complicated by the fact that **1** (*e.g.* as purchased from *Sigma Aldrich*) exists as a racemate of the enantiomers *(4aS,8aR)*-**1** and *(4aR,8aS)*-**1** (henceforth *SR*-**1** and *RS*-**1**, and with dark *vs.* light blue color codes, respectively, in Chart 1 and subsequent figures). Consequently, reduction of **1** can yield four different stereoisomers of **2** (Chart 1, *center*): when attacked by H^- in *pro*-*R* orientation, *SR*-**1** can only yield *(1R,4aS,8aR)*-**2** (henceforth *RSR*-**2**; black in Chart 1), and *RS*-**1** can only yield *(1R,4aR,8aS)*-**2** (*RRS*-**2**; grey); conversely, when attacked in *pro*-*S* orientation, *SR*-**1** and *RS*-**1** can only yield *(1S,4aS,8aR)*-**2** and *(1S,4aR,8aS)*-**2**, respectively (*SSR*-**2** and *SRS*-**2**; red and orange in Chart 1). Thus, when simulating the enzymatic reaction to study its stereoselectivity in the "forwards" direction, all four outcomes need to be considered. For illustration, putative Michaelis complexes in WT *act*KR for these scenarios are shown in Figure 2.

Due to increases in computer speed and improvements in algorithms, parameters, and usability over the past decades,[18] *in silico* biomolecular simulation is increasingly used to complement experimental enzymatic studies,[19-21] including those focusing on the design of novel enzyme variants relevant for biocatalysis.[22-25] Indeed, enzyme reactivity and stereocontrol are often the result of a subtle interplay between different factors that are difficult to examine separately experimentally.[26] In short, enzyme variants conferring stereoselectivity can either do so by favoring formation of a particular reactive complex; or by favoring efficient reaction of the stereochemistry-determining step (or subsequent steps); or a combination of both. Here, we assess the stereoselectivity of *act*KR variants using three different factors that can easily be assessed individually through simulation.

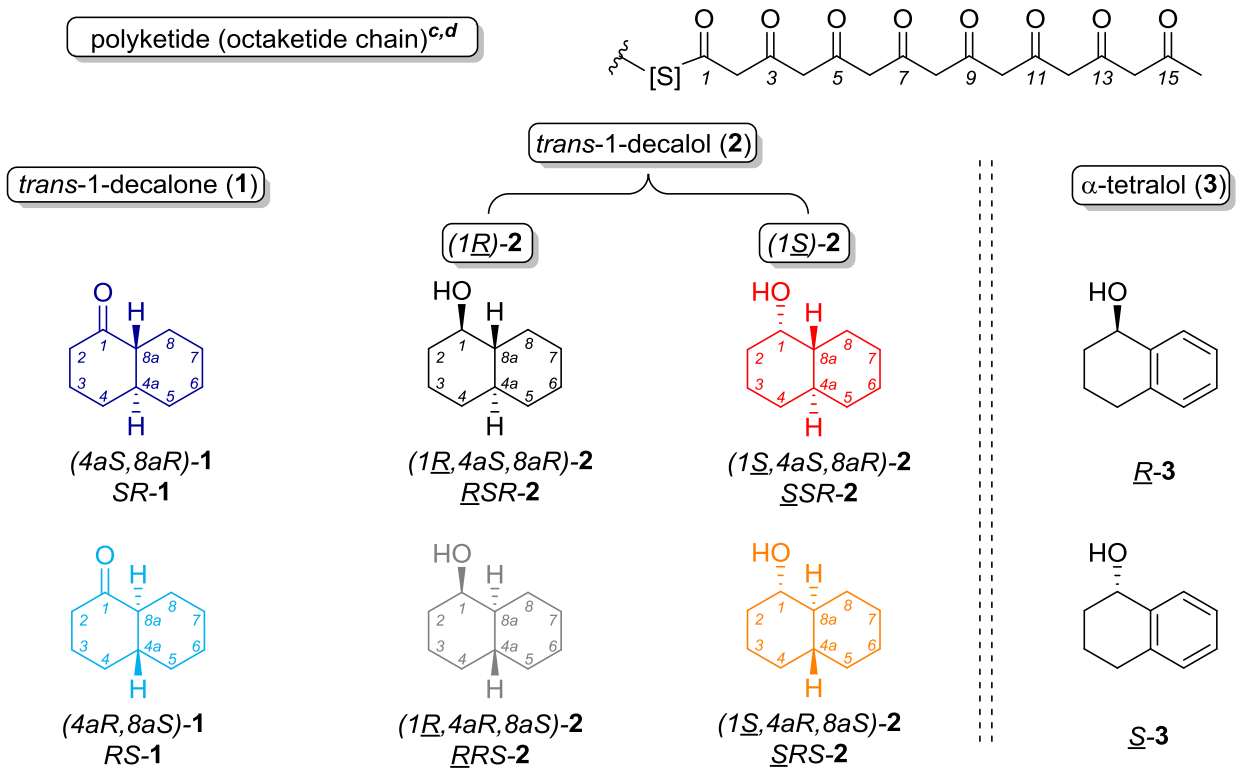
Factor I (henceforth **FI**) is the likelihood of the enzyme-substrate complex attaining a "reactive" pose (or Michaelis complex); or, in other words, how likely it is for the substrate to reach an orientation and conformation at which key interatomic distances are sufficiently close for a reaction to occur. Any imbalance of this factor in favor of reactive *pro*-*R* or *pro*-*S* orientations will thus affect an enzyme's stereocontrol by giving a head start to one reaction pathway over the other. In principle, **F1** describes the energetics of reaching and maintaining a reactive pose.[27] This incorporates the interplay between conformational fluctuations in the enzyme and substrate ('dynamics'), and may even be affected by mutations far from the active site.[26] Computationally, this can be probed by classical molecular dynamics (MD) simulations; in some cases (*e.g.* with sufficient structural information and particular prochiral binding orientations precluded), molecular docking may suffice.[28-29]

Factor II (henceforth **FII**) can be seen as part of **FI**, but focusing solely on the end point, *i.e.* the binding affinity (or binding free energy; ΔG_{bind}) of a specific substrate in its reactive pose within an enzyme's active site. Assuming that reactive poses can be attained by both prochiral orientations, one prochiral orientation may be preferred over the other (i.e. $\Delta\Delta G_{bind} \neq 0$), thereby favoring formation of the product associated with the preferred prochirality. *In silico* methods for calculating ΔG_{bind} include (but are by no means limited to) *WaterSwap*,[30] enhanced Monte Carlo,[31] and MM/PBSA.[32-35]

Finally, Factor III (henceforth **FIII**) is the height of the free energy barrier separating reactants and products of different chirality: if a substrate in a reactive *pro*-*R* or *pro*-*S* orientation is closer in energy to the corresponding transition state, then the associated product will be more accessible than its counterpart. A range of computational approaches are available to determine such barriers in enzyme reactions;[36-38] a popular option is the use of a hybrid quantum / classical (QM/MM) approach,[39] where the computationally costly QM treatment can be limited to regions of chemical change (without time-intensive parameterization).[37-39]

In the present work, we describe our computational efforts to rationalize the stereoselectivity of WT, P94L, and V151L *act*KR towards **1**, determining the role of **FI**-**FIII** in each case. Based on the assumption that the stereospecific oxidation of **3** is comparable to the stereoselective reduction of **1**, we expect P94L *act*KR to preferentially form *(1S)*-*trans*-1-decalol; V151L *act*KR to preferentially form *(1R)*-*trans*-1-decalol; and a slight preference for *(1S)*-*trans*-1-decalol by WT *act*KR. Focusing on a potential biocatalyst that has thus far been little studied *in silico*, we aim to show that physically realistic but relatively inexpensive computational simulations can be readily employed in this and similar contexts, to explain the behavior of other promising biocatalysts, and helping to direct the design of novel mutants with enhanced stereocontrol.

Chart 1. *act*KR substrates discussed in this work.[a,b]



[a] **1** and **2** color-coded as in Figure 4 and Figure 6, respectively. [b] Underlined R or S labels in **2** and **3** (and throughout the text) mark stereocenters introduced/removed by *act*KR. [c] C7-C12 cyclisation prior to reduction at C9. [d] [S] denotes link to ACP (see text).

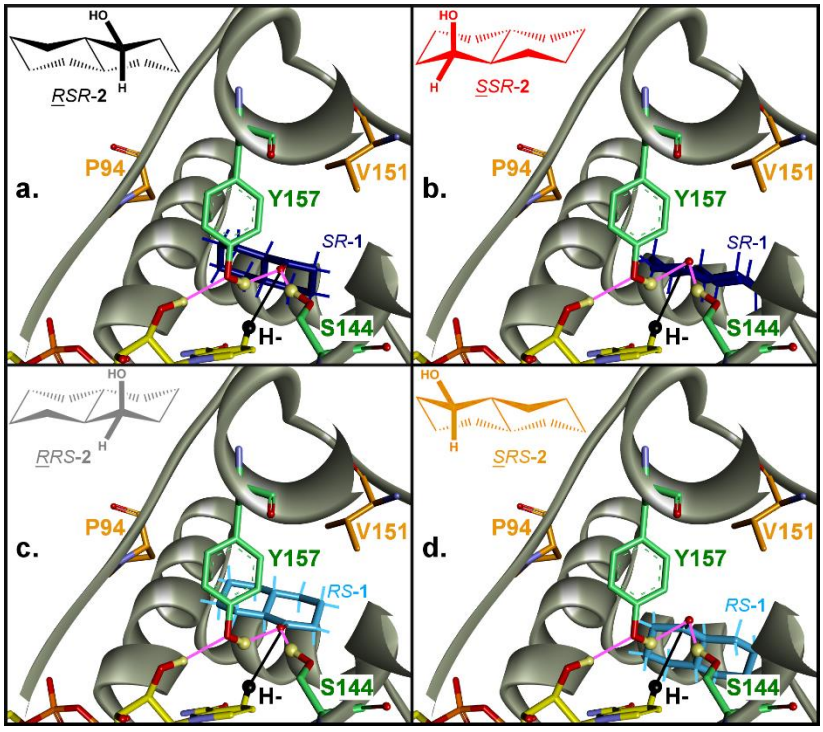


Figure 2. Comparison of (*a*, *b*) *pro*-R and *pro*-S poses of *SR*-**1** in WT *act*KR; and (*c*, *d*) as before, but with *RS*-**1**. In the top left of each panel, we indicate the enantiomer of **2** generated in each case (*cf.* Chart 1). Snapshots are taken from the opposite angle to that in Figure 1b, and salient protons are shown in light yellow; colors and labels are otherwise identical.

2. Computational Strategy and Details

General Procedure. For **FI**, **FII**, and **FIII** investigations alike, the first step entails running a number of independent classical MD simulations, to sample conformational space at relatively low computational cost.

In MD simulations for **FI**, which we henceforth refer to as "free", the substrate is left free to explore as many orientations as possible in the binding site: the frequency of reactive *pro*-*R* and *pro*-*S* poses may be directly extracted from these runs. To avoid substrate diffusion from the active site, a one-sided harmonic restraint (k = 50 kcal mol^{-1} Å^{-2}) is employed when the center of mass of the substrate moves farther than 8 Å from the center of mass of NADPH's nicotinamide moiety.

To provide starting points for evaluating **FII** and **FIII**, MD simulations were also performed of the "Michaelis" or "reactive" *pro*-*R* and *pro*-*S* enzyme-substrate complexes, from which, in principle, the chemical reaction can readily begin. These "restrained" MD simulations include one-sided harmonic restraints on three key interatomic distances, and an additional dihedral angle restraint to prevent the substrate from 'flipping' between *pro*-*R* and *pro*-*S* (see *Supporting Information* for further details).

To determine the role of **FII**, ΔG_{bind} of *SR*-**1** and *RS*-**1** in their two prochiral orientations is calculated through a series of MM/PBSA calculations[32-35] (*vide infra*) on the resulting "restrained" trajectories. Probing the role of **FIII** requires adequate sampling of free energy barriers for the conversion of **1**'s two enantiomers to one of the four accessible stereoisomers of **2**. To do this, we select a total of 120 representative snapshots from the "restrained" MD simulations and carry out, on each snapshot, hybrid quantum/classical MD (QM/MM MD) with umbrella sampling (US) along a reaction coordinate (*vide infra*).[38]

For simplicity, all classical and QM/MM MD simulations are run in two distinct sets, treating enantiomers *SR*-**1** and *RS*-**1** as separate substrates and, thus, with only one or the other enantiomer occupying all four active sites in each set.

Starting Structures. Six starting structures [*act*KR-(NADPH)$_4$-**1**$_4$] are set up with all four sites occupied: *SR*-**1** or *RS*-**1** in WT, P94L, or V151L *act*KR. All starting structures are constructed from our reference WT crystal structure (PDB ID: 2RH4),[14] to demonstrate the possibility of assessing enzyme variants for which no crystal structure has been obtained. Existing mutant crystal structures (V151L; PDB ID: 4DBZ[17] and P94L; PDB IDs: 3RI3, 3QRW)[16] are checked to ensure that constructed mutants retain plausibly oriented sidechains. Constructed structures are henceforth labelled as follows: WT-*SR*-**1**$_4$ (*cf.* Figure 2a and 2b), V151L-*SR*-**1**$_4$, and P94L-*SR*-**1**$_4$ for one enantiomer of **1**; and WT-*RS*-**1**$_4$ (*cf.* Figure 2c and 2d), V151L-*RS*-**1**$_4$, and P94L-*RS*-**1**$_4$ for the other.

In all six cases, active sites in monomers A and B are populated with the substrate in *pro*-*R* orientation whereas those in monomers C and D are populated with the substrate in *pro*-*S* orientation. All four monomers are modelled from residues 1 to 261. All residues are modelled in their standard protonation states, in line with pK_as predicted by PROPKA 3.1.[40] Hydrogens are added with *AmberTools'* (version 17)[41] *reduce* utility, resulting in His162 being singly protonated on Nδ1; His153 and His201 on Nε2. Using *tleap*,[41] standard *N*- and *C*-termini are introduced and the structure is solvated in a truncated octahedral box of water extending at least 11 Å from any protein atom, with 40 Na^+ ions added to neutralize the system. Starting structure files and further details regarding starting structure generation are available as *Supporting Information*.

Classical MD. MD simulations of our six *act*KR-(NADPH)$_4$-**1**$_4$ systems in explicit water are run with the *AMBER* software package (version 16, 2017 distribution),[41-42] taking advantage of GPU acceleration where applicable.[43] Postprocessing and analysis of MD trajectories are carried out with the *CPPTRAJ* utility;[44] and visual inspection is conducted with *VMD*.[45] The protein and ions are described by the *ff14SB* forcefield,[46] the NADPH cofactor by the forcefield from Holmberg and coworkers,[47] and water using the TIP3P model.[48] The *GAFF* forcefield[49] with AM1-BCC charges derived by *antechamber*[50-51] is used for *SR*-**1** and *RS*-**1** (see also *Supporting Information*). The default cutoff of 8 Å is used to compute Lennard-Jones and Coulomb interactions, with Coulomb interactions beyond this limit computed using the Particle Mesh Ewald method.[52-53]

For both the "free" and "restrained" MD simulations (*vide supra*) we carry out eight independent simulations (different random seeds) for each of the six *act*KR-(NADPH)$_4$-**1**$_4$ starting structures. This results in a total of 8 × 6 = 48 MD runs for both the "free" and "restrained" sets (96 in total). Each of the MD runs is carried out with the following general procedure (restraints retained throughout): minimization (600 steps); solvent equilibration (9 ps, *NVT*); heating (20 ps, *NVT*); equilibration (2040 ps, *NpT*); and production (12 ns, *NpT*). Production runs are conducted in the *NpT* ensemble with a time-step of 2 fs. A constant pressure of 1 atm is enforced *via* the Berendsen barostat[54] and a constant temperature of 298 K is enforced *via* the Langevin thermostat[55] (collision frequency set at 5 ps^{-1}). Bonds containing hydrogen are constrained by employing the *SETTLE* and *SHAKE* algorithms.[56] (See *Supporting Information* for details on settings in the pre-production stages).

MM/PBSA Calculations. MM/PBSA calculations[32-35] are run on 500 snapshots from each restrained MD trajectory, on 16 processors, using *AmberTools*' *MMPBSA.py.MPI* utility,[57] with default atomic radii,[58] settings, and parameters.[42, 57] The only exceptions are the ionic strength, which is set at 0.025 mol dm^{-3} to reflect the 40 Na^+ ions originally present in the simulations; and the internal dielectric constant ε_{int}, which is set to 4.0 as advised by Wang and coworkers.[35] The six *act*KR-(NADPH)$_4$-**1**$_4$ topologies are preprocessed using the *ante-MMPBSA.py* tool.[57]

The eight trajectories of each of the six *act*KR-(NADPH)$_4$-**1**$_4$ systems are parsed by *MMPBSA.py.MPI* as six individual 96 ns "supertrajectories", and each MM/PBSA calculation then runs on snapshots taken every 24 ps. Since each system contains four instances of **1** (*i.e.* one in each active site), four separate MM/PBSA runs are required on each "supertrajectory" (*i.e.* 24 in total, giving 24 ΔG_{bind} values). Entropic corrections to these values are calculated using the interaction entropy method reported by Duan *et al*.[59]

QM/MM MD with Umbrella Sampling. All QM/MM MD simulations are run with *AMBER*'s *sander* MD engine.[60-61] Apart from the time-step (here 1 fs) and the absence of restraints other than the reaction coordinate, all other conditions

are identical to the classical MD production runs. The QM region (with explicit treatment of electrons) comprises: the reacting substrate in its entirety; NADPH's nicotinamide moiety up to the first ribose; the Ser144 and Tyr157 sidechains (from C_β); and three hydrogen link-atoms at the QM-MM boundary (Figure 3). All remaining atoms are included in the classical (MM) region. The *SHAKE* algorithm is switched off in the QM region, and atoms are treated using the semiempirical method PM6.[62] PM6 was chosen over other semiempirical methods after benchmarking the reaction on a small model with higher-level QM calculations (SCS-MP2[63-64] single-point energy calculations on B3LYP[65]/6-31+G(*d*,*p*)-optimized potential energy surfaces); PM6 correctly identified the mechanism with a highly similar transition state, even though it significantly overestimates the barrier for reaction and the reaction energy. Details of benchmarking calculations are provided as *Supporting Information*; relevant coordinates and frequency calculations are also available[66] for all stationary points in the *ioChem-BD* repository.[67]

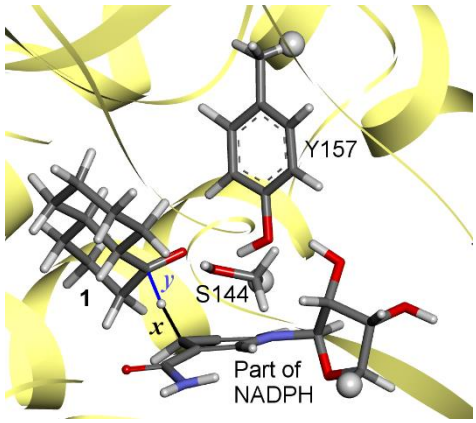


Figure 3. QM region for QM/MM MD US simulations (rendered as sticks, except for three capping H atoms rendered as white spheres). Key: white: H; grey: C; blue: N; red: O. Secondary structure of MM region rendered as yellow ribbon. The chosen reaction coordinate (x – y) is marked with blue and black lines; atoms included in it are defined in the text.

Simulations are started from a series of representative reactive snapshots (see *Supporting Information*) from the restrained MD simulations, with retention of atom velocities. Ten such snapshots are chosen for each of the twelve reaction paths to be compared: reduction of *SR*-**1** to either (1) *R*,*SR*-**2** or (2) *S*,*SR*-**2**; and reduction of *RS*-**1** to either (3) *R*,*RS*-**2** or (4) *S*,*RS*-**2**; in WT; (5-8) P94L; and (9-12) V151L *act*KR. Only one of the four substrates present in each snapshot is considered for umbrella sampling (US).

Hydride transfer (*i.e.* the reaction coordinate) is monitored using the difference between two key distances (Figure 3): the distance NADPH:H^-–**1**:C1 (y) is subtracted from the distance NADPH:H^-–NADPH:C_{H-} (x). Umbrella sampling (US) is performed by lengthening a restraint on the reaction coordinate (x – y) in 0.1 Å steps ("windows") starting from a value close to that found in the initial snapshot, and thus gradually "forcing" the system to react until (x – y) reaches a value of 1.8 Å. Each US window is simulated with QM/MM MD for 2 ps, and (x – y) is restrained at the desired length through a harmonic biasing potential (k = 100 kcal mol^{-1} Å^{-2}): the last configuration of each US window is used to start the QM/MM MD simulation for the following one. Depending on the starting value of (x – y), the simulation time for each individual US simulation ranges from 66 ps (33 windows) to 90 ps (45 windows), with the total cumulative simulation time for each free energy profile ranging from 728 ps to 840 ps. The potential of mean force (PMF) along (x – y) amounts to the free energy profile of the reduction, and was obtained from the US runs using the weighted histogram analysis method (WHAM)[68-69] with the eponymous program by Grossfield:[70] this also carries out error analysis through Monte-Carlo bootstrapping. PMFs are extracted both individually for all 120 US runs, and cumulatively for the 12 processes sampled. A number of runs were discarded as outliers, for example when they did not manage to capture the concerted proton abstraction as observed in our benchmarking studies; all discarded runs were replaced. Details regarding reruns, WHAM, and error analysis are given in the *Supporting Information*.

3. Results and Discussion

In the subsections below, we first discuss findings from our calculations regarding the individual effects of **FI**-**FIII** on *act*KR stereoselectivity upon reduction of **1** to **2**. We then examine which combinations of factors come into play when determining stereocontrol in WT, P94L, and V151L *act*KR, drawing comparisons with corresponding *in vitro* observations on the oxidation of **3**,[14, 16-17] and including trends between enantiomers of **1**. Finally, we discuss our protocols and their potential for high-throughput screening, focusing on the simulation lengths required to achieve appreciable accuracy.

Formation of Reactive Complexes: FI. We measure the difference (Δ) between the frequency of reactive *pro*-*R* and reactive *pro*-*S* poses of *SR*-**1** ($\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*SR*-**1**) and *RS*-**1** ($\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*RS*-**1**) in WT, P94L, and V151L *act*KR as they occur in our "free" MD simulations. Criteria to determine whether or not a pose is "reactive", and whether its prochirality is *pro*-*R* or *pro*-*S*, are based on the restraints imposed on "restrained" MD simulations used to study **FII** and **FIII** (*Supporting Information*, Table S1).

For each of our six simulated systems (WT-*SR*-$\mathbf{1}_4$, V151L-*SR*-$\mathbf{1}_4$, P94L-*SR*-$\mathbf{1}_4$, WT-*RS*-$\mathbf{1}_4$, V151L-*RS*-$\mathbf{1}_4$, and P94L-*RS*-$\mathbf{1}_4$), statistics for $\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*SR*-**1** or $\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*RS*-**1** are measured cumulatively: in other words, the 8 MD replicas conducted for each of the six systems are considered collectively, as are all 4 active sites. $\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*SR*-**1** and $\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*RS*-**1** are calculated for each variant using:

$$\Delta_{[pro\text{-}R-pro\text{-}S]}SR\text{-}\mathbf{1} = \left(\frac{\text{reactive } pro\text{-}R\ SR\text{-}\mathbf{1} \text{ poses}}{96000} - \frac{\text{reactive } pro\text{-}S\ SR\text{-}\mathbf{1} \text{ poses}}{96000}\right) \times 100$$

and

$$\Delta_{[pro\text{-}R-pro\text{-}S]}RS\text{-}\mathbf{1} = \left(\frac{\text{reactive } pro\text{-}R\ RS\text{-}\mathbf{1} \text{ poses}}{96000} - \frac{\text{reactive } pro\text{-}S\ RS\text{-}\mathbf{1} \text{ poses}}{96000}\right) \times 100\ ;$$

where the denominator 96000 reflects the fact that each of the eight replicas we are dealing with has 3000 frames (250 per ns), and that each of these frames has 4 active sites (and 4 instances of **1**). A general Δ for *trans*-1-decalone as a whole ($\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$**1**) is then obtained by summing $\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*SR*-**1** and $\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}$*RS*-**1**.

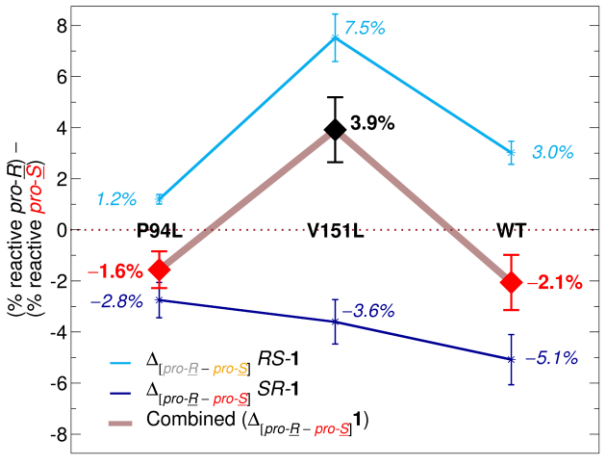


Figure 4. Difference (Δ) between the frequency of reactive *pro*-R and reactive *pro*-S poses in "free" MD simulations of: (*dark blue*) *SR*-**1**; (*light blue*) *RS*-**1**; and (*brown*) **1** overall, within P94L, V151L, and WT *act*KR. Negative values indicate excess of *pro*-S reactive poses (red points), positive values excess of *pro*-R poses (black point). See also color codes in Chart 1; lines are to guide the eye. Errors indicated are based on leave-one-out procedure (details in *Supporting Information*).

Although the overall frequency of *pro*-R and *pro*-S reactive poses is generally low (the highest being 8.3% across the 8 V151L-*RS*-$\mathbf{1}_4$ MD replicas), significant trends still emerge (Figure 4). The first observation to make is that, regardless of *act*KR variant, simulations with *SR*-**1** (dark blue) consistently show an excess of reactive *pro*-S poses (most marked in WT-*SR*-$\mathbf{1}_4$ at 5.1%), whereas *RS*-**1** systems (light blue) all favor reactive *pro*-R poses instead (most markedly in V151L-*RS*-$\mathbf{1}_4$ at 7.5%).

A second noteworthy observation is that for the overall difference in pro-R / pro-S reactive poses ($\Delta_{[pro\text{-}R\ -\ pro\text{-}S]}\mathbf{1}$; brown line in Figure 4), V151L *act*KR has an excess of reactive *pro*-R poses (3.9%; black diamond in Figure 4), whereas both WT and P94L *act*KR have an excess of reactive *pro*-S poses (red diamonds).

The excess of reactive *pro*-R poses seen in V151L *act*KR is encouragingly in line with the mutant's observed *in vitro* specificity towards R-**3** over S-**3** in the reverse direction.[17] Similarly, the slight excess of *pro*-S reactive poses emerging from simulations of P94L and WT *act*KR is in line with *in vitro* observations that both are preferentially turning over S-**3**.[16] Nonetheless, if **FI** were the predominant factor driving P94L *act*KR's exclusive preference for S-**3**, one would expect a far greater excess of reactive *pro*-S poses to emerge.

Binding Free Energy of Reactive Poses: FII. The role of **FII** was investigated through separate MM/PBSA binding free energy calculations on each of the four active sites of our six *act*KR-$(NADPH)_4$-$\mathbf{1}_4$ systems (using snapshots taken at 24 ps-intervals from 8 independent restrained MD runs of 12 ns each). ΔG_{bind} values for active sites A and B (containing *pro*-R poses) and those for active sites C and D (containing *pro*-S poses) are averaged and plotted in (Figure 5).

For *SR*-**1**, ΔG_{bind} values for *pro*-R and *pro*-S poses are highly similar (left-hand side of Figure 5), indicating no particular preference for either prochirality in any of the enzyme variants. The only significant difference is that the P94L mutant indiscriminately shows a higher affinity for both orientations (more negative by ~1.5 kcal mol^{-1} with respect to WT and V151L *act*KR). More variation is observed between the binding energies for *pro*-R and *pro*-S poses of *RS*-**1** (right-hand side of Figure 5). Whilst both orientations retain similar binding affinities in WT *act*KR ($\Delta\Delta G_{bind} \approx 0$), this is clearly not the case in the two mutants: *RS*-**1** in *pro*-R orientation has a significantly higher affinity (compared to its *pro*-S orientation) in both V151L *act*KR ($\Delta\Delta G_{bind}$ = 1.9 kcal mol^{-1}) and P94L *act*KR ($\Delta\Delta G_{bind}$ = 3.6 kcal mol^{-1}). This increase in $\Delta\Delta G_{bind}$ when going from V151L to P94L *act*KR is mainly caused by a loss in affinity for the *pro*-S orientation.

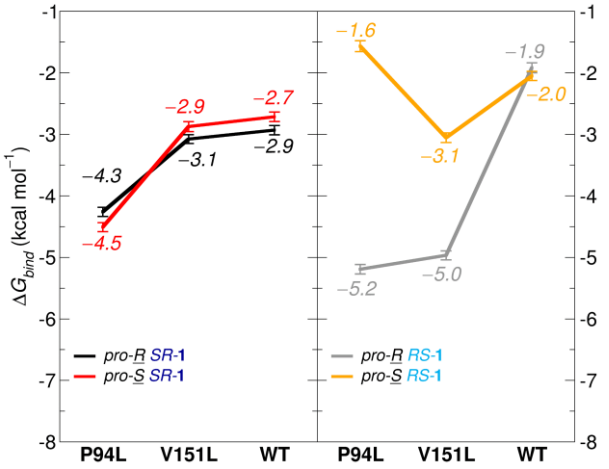


Figure 5. Binding free energies (ΔG_{bind}) predicted by MM/PBSA calculations on (*left*) *SR*-**1**; and (*right*) *RS*-**1**; either in *pro*-R poses (black or grey, respectively); or in *pro*-S poses (red or orange, respectively). Observations for *pro*-R poses are averaged between monomers A and B; those for *pro*-S between C and D. See also color codes in Chart 1; lines are guides for the eyes. Error bars correspond to the standard error of the mean.

In the case of WT *act*KR, enantiomers of **1** all have similar binding affinities regardless of orientation (with a slight preference of *SR*-**1** in either pose over *RS*-**1**, $\Delta\Delta G_{bind} \approx 0.75$ kcal mol^{-1}). In the case of V151L, only the *pro*-R *RS*-**1** pose has significantly higher affinity, with *pro*-S *RS*-**1** at the same level as both *SR*-**1** orientations (in line with the preferential formation of reactive poses for *pro*-R *RS*-**1**, *vide supra*). For P94L, there is a significant loss in affinity of *pro*-S *RS*-**1** (orange), whereas *pro*-R *RS*-**1** and both *SR*-**1** poses have similar affinities (within 1 kcal mol^{-1}). Based on the stereoselectivity observed *in vitro*,[14, 16-17] one might expect to see a greater (more negative) binding affinity of *pro*-S poses for at least one of *SR*-**1** / *RS*-**1** (red / orange) in the case of P94L (as well as a greater binding affinity of *pro*-R poses in at least one of *SR*-**1** / *RS*-**1** in the case of V151L, as is observed here). Thus, **FII** is not likely to be involved in determining the stereoselectivity of P94L *act*KR, but may contribute instead to that of V151L *act*KR.

Free Energy Barrier of Reaction: FIII. To probe the effects of **FIII** on the stereoselectivity of WT *act*KR and its mutants, we have used QM/MM MD umbrella sampling simulations to obtain free energy profiles for each of the four possible reductions of *trans*-1-decalone to *trans*-1-decalol (**1** to **2**, Figure 6). Profiles have been normalized so that the free energy of **1** on the "reactants" side is equal to zero, and each plot has been color-coded according to the stereoisomer of **2** obtained (*cf.* color codes in Chart 1). Note that the unusually high energy barriers and reaction energies are mainly due to

the use of PM6 for the QM region (see *Supporting Information*, Figure S2); in addition, the small QM region, the approximate reaction coordinate and, in case of reaction energies, the fact that we do not sample subsequent proton transfers,[15] will likely contribute to overestimation. We note that *differences* in reaction barrier height are what is relevant for investigation of the influence of **FIII** on stereoselectivity.

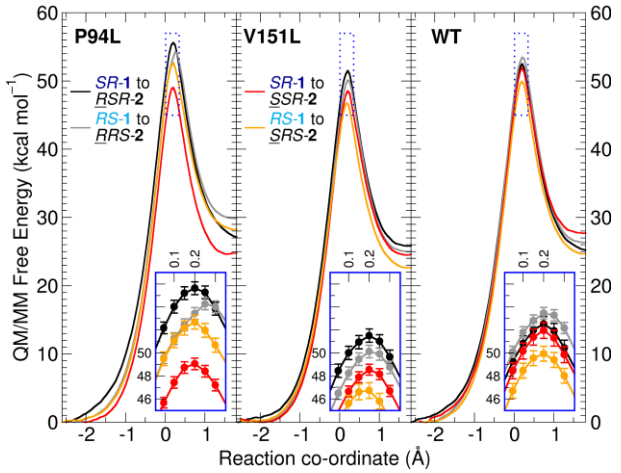


Figure 6. Free energy profiles resulting from QM/MM MD US simulations (PM6/ff14SB) of the conversion of **1** to **2** in: (*left*) P94L *act*KR; (*center*) V151L *act*KR; and (*right*) WT *act*KR. In each panel, each curve is the cumulative result of 10 individual simulations run for the 4 possible combinations of: *SR*-**1** or *RS*-**1** (darker- *vs.* lighter-colored curves); and final *R*-**2** or *S*-**2** stereochemistry (black/grey *vs.* red/orange, respectively; *cf.* Chart 1). Regions close to the transition state (blue dotted line) are magnified in the blue insets, with error bars; errors are obtained through Monte-Carlo bootstrapping (see *Supporting Information* for details).

For P94L *act*KR (Figure 6, *left* panel), *SSR*-**2** (red) is indicated as the most easily accessible stereoisomer of **2**, through a free energy barrier that is at least 5 kcal mol^{-1} lower than those required to access either *R*-**2** stereoisomer, and about 4 kcal mol^{-1} lower than that required for the formation of *SRS*-**2**. This finding is in line with the *S*-specificity detected *in vitro*,[16] and is further substantiated by the fact that reaching either of the two *R*-**2** stereoisomers (black and grey) is far more difficult in this variant than in either V151L or WT *act*KR. The fact that the barrier to reach the other *S*-**2** stereoisomer (*SRS*-**2**; orange) is also higher than in the other forms of *act*KR—together with the aforementioned lower binding affinity for the *pro*-*S* *RS*-**1** reactive pose (Figure 5)—suggests that *S*-**2** selectivity in P94L *act*KR is likely to arise from formation of *SSR*-**2** from *SR*-**1**.

Findings for V151L (Figure 6, *center* panel) are less clear: although barriers to reach the two *R*-**2** stereoisomers are indeed significantly lower than in P94L and WT *act*KR, barriers to reach the *S*-**2** stereoisomers are both even lower, with formation of *SRS*-**2** exhibiting the lowest free energy barrier out of all investigated combinations. Taken on its own, this contradicts the *in vitro* observation (specificity towards *R*-**3** *vs.* *S*-**3**):[17] other factors are thus likely to play a more dominant role (*vide infra*).

For WT *act*KR (Figure 6, *right* panel), a slightly lower free energy barrier is only detected for the conversion of *RS*-**1** to *SRS*-**2** (orange; 2.1 to 3.6 kcal mol^{-1} lower than the other three). None of the other three conversions stand out for being particularly "easier". This may agree with only a mild preference for *S*-**3** being observed experimentally.[14, 16] Finally, we note that the reaction barriers for the expected preferred reactions are similar between the three enzyme variants. This is consistent with experiment,[4, 14, 17] considering our approximate reaction simulation approach: the P94L and V151L mutations only have very minor effects on the measured conversion rate (k_{cat}) of racemic **1** to **2**, which translate (using transition state theory) to subtle increases in the free energy of activation ($\Delta G^{\ddagger}$) of 0.8 kcal mol^{-1} and 0.4 kcal mol^{-1}, respectively.

Different Factors Dominate for Different Variants. Taking into account the data for **FI**-**FIII** arising from our different simulations (Figure 4, Figure 5, and Figure 6), it becomes clear that the causes driving stereocontrol in each *act*KR variant are likely to be different.

For V151L *act*KR, expected to have a strong *R*-preference, the *RS*-**1** enantiomer is significantly more prone to form reactive *pro*-*R* poses in the enzyme active site (**FI**). In addition, binding of reactive *pro*-*R* *RS*-**1** poses is thermodynamically favored over the other possibilities (**FII**). It is true that reaching *RRS*-**2** from *pro*-*R* *RS*-**1** is significantly easier in V151L *act*KR than in any of the other variants investigated (**FIII**; Figure 6), however, this in itself this would not enhance stereoselectivity, as barriers to other species are equally reduced. It therefore appears that for V151L *act*KR, stereoselectivity is mainly determined by the sterics and thermodynamics prior to the reduction step (**FI** and **FII**). As such, our calculations indicate that this variant will predominantly form *RRS*-**2** (instead of *RSR*-**2**).

In the case of the WT enzyme, mild selectivity can be expected towards *S*-**2** (based on its experimentally determined specificity towards *S*-**3**). The *pro*-*S* reactive poses are somewhat more easily attained within its active site (**FI**; Figure 4), with *SR*-**1** being the best-placed enantiomer to reach its *SSR*-**2** product. Thus, for *SSR*-**2** formation (compared to *RSR*-**2**), this "steric head start" is likely important. Formation of the other *S*-**2** isomer, *SRS*-**2**, is favored by its lower free energy barrier (**FIII**; Figure 6). **FII** (Figure 5) appears to play no significant role for stereoselectivity.

For P94L *act*KR, the situation is significantly different: its strong *S*-**2** selectivity (as expected from its *in vitro* specificity towards *S*-**3**) mostly arises due to a change in relative reaction barriers (**FIII**). Specifically, formation of *SSR*-**2** is clearly preferred (Figure 6). **FI** may play some additional role, since *SR*-**1** forms more *pro*-*S* reactive poses. Altogether, the simulations further indicate that this variant will predominantly form *SSR*-**2** (instead of *SRS*-**2**): *pro*-*S* *RS*-**1** reactive poses occur less frequently and have lower affinity than *pro*-*S* *SR*-**1** reactive poses (Figure 4; Figure 5), and they are also much less favored to react to *SRS*-**2** (**FIII**; Figure 6).

Table 2. Key Structural Features of Reactive Poses of 1

Enantiomer	*SR*-**1**	*SR*-**1**	*RS*-**1**	*RS*-**1**
Prochirality	*pro*-*R*	*pro*-*S*	*pro*-*R*	*pro*-*S*
***Non*-CO ring**[a]	→ 94	→ 151	→ 94	→ 151
H$^-$ in 2[b]	Axial	Equatorial	Equatorial	Axial

[a] Approximate direction towards which **1**'s C4a-C5-C6-C7-C8-C8a ring points: P/L94 *vs.* V/L151. [b] Final position of the transferred hydride in the resulting isomer of **2**.

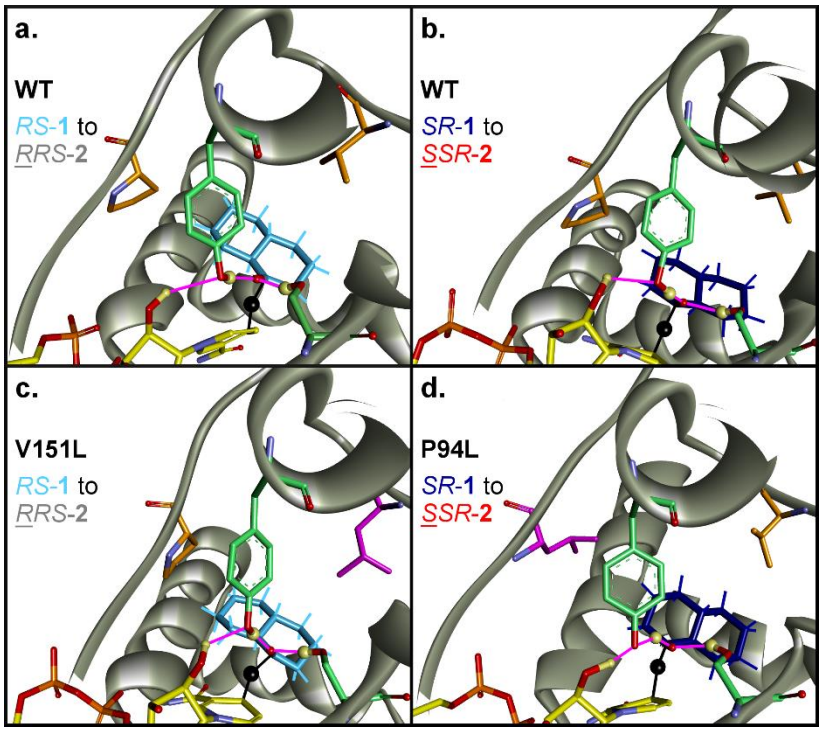


Figure 7. Representative snapshots of approximate transition states from QM/MM MD US simulations (obtained from clustering on the substrate RMSD after alignment on the cofactor). (*a*) *pro*-R *RS*-**1** to R*RS*-**2** in WT *act*KR; (*b*) *pro*-S *SR*-**1** to S*SR*-**2** in WT *act*KR; (*c*) *pro*-R *RS*-**1** to R*RS*-**2** in V151L *act*KR; and (*d*) *pro*-S *SR*-**1** to S*SR*-**2** in P94L *act*KR. In each case, snapshots are taken from the 0.2 simulation window; all four choices depict equatorial attack. Key and orientation: same as Figure 2; mutations in (*c*) and (*d*) are marked in magenta.

Our extensive atomistic simulations allow investigation of geometrical characteristics within each active site that may help rationalize the above findings. Considering the orientation of *RS*-**1** and *SR*-**1** in the active site (Table 2); as well as representative transition states for their reaction to **2** (Figure 7), it emerges that (pro)chiralities enhanced by each mutation (R for V151L; S for P94L) are those which do not entail a steric clash between the *non*-carbonyl ring of **1** and the extra bulk introduced by that mutation. We note that, in this case, this effect could not be reproduced by standard molecular docking (likely due to the active site's large volume compared to **1**).

There appears to be a slight difference in the effect brought about by mutant sidechains. The L151 sidechain (Figure 7c) extends more directly into the active site, and thus influences prochirality by 'stealing' volume from prospective *pro*-S poses (from the side of the non-carbonyl ring), making such poses less likely (**FI**). The L94 sidechain appears to have a subtler effect (Figure 7d), potentially affecting the orientation of the substrate during reaction (thereby affecting **FIII**).

Trans-1-decalone Enantiomers and Equatorial Attack. Our work indicates that P94L *act*KR prefers facilitating the reaction of *pro*-S *SR*-**1**, and V151L *act*KR prefers facilitating *pro*-R *RS*-**1**. A significant difference in *SR*-**1** and *RS*-**1** complexes with *act*KR already emerged from investigation of **FI** (Figure 4), indicating a general preference of *pro*-S reactive poses with *SR*-**1** (dark blue), and *pro*-R reactive poses with *RS*-**1** (light blue), regardless of the *act*KR variant considered. These observations, in combination with the geometrical features from the simulations (Table 2 and Figure 7), indicate that both enantiomers of **1** are prone to react in the prochirality that favors *equatorial* rather than axial hydride attack. This contrasts with previous literature on the non-enzymatic reduction of **1** *in vitro*[71-72] (with reagents such as $[BH_4]^-$ and $[AlH_4]^-$), which shows that the favored product is instead the one resulting from axial attack. A similar preference for the axial product was found computationally for the non-enzymatic reduction of cyclohexanone,[72] with the transition state for axial attack found to be about 1.8 kcal mol^{-1} more stable.

The preference we predict for equatorial hydride attack in the enzymatic reduction by *act*KR is, however, entirely in line with experimental findings by Østergaard *et al.*,[73] which showed *SR*-**1** (from a racemate of **1**) to preferentially undergo equatorial attack (to form S*SR*-**2**) within the ketoreductase module of the erythromycin polyketide synthase. The authors further concluded that, in the reverse direction, the enzyme preferentially turns over a racemate of R*RS*-**2** and S*SR*-**2** (as opposed to one of R*SR*-**2** and S*RS*-**2**)

Does Selectivity for Trans-1-decalone Reduction Equal Specificity for α-tetralol Oxidation? The significant difference between *act*KR complexes with *RS*-**1** and *SR*-**1** prompts a deeper discussion on the general assumption that selectivity towards *trans*-1-decalone and *α*-tetralone are directly comparable. Resolving the stereochemical outcome of the "forwards" reaction with *trans*-1-decalone (by assessing S-**2** : R-**2** ratio) was not successful, despite trying several approaches (*personal communication* with Prof. SC Tsai, UC Irvine). Korman, Javidpour *et al.*[14, 16-17] thus investigate specificity in the reverse reaction for S-**3** and R-**3** (commercially available, unlike the stereoisomers of **2**); studying the "forwards" reduction of *α*-tetralone is hampered by *act*KR's very sluggish turnover of that substrate.

To obtain detailed insight into (the stereoselectivity of) reduction of *trans*-1-decalone, which is readily turned over by the three enzyme variants studied here (Table 1), we opted to

investigate the "forwards" reaction (reduction of **1** to **2**). Note that had we chosen to study the reaction using *α*-tetralone/**3** instead, a complication would have arisen because the aliphatic carbonyl ring can chair-flip freely (due to aromaticity in its non-carbonyl ring): chair-flipping—virtually unachievable in **1**—would have introduced a considerable degree of complexity in our simulations. We thus rely on the assumption (also made by Korman, Javidpour *et al.*)[14, 16-17] that the reduction of **1** to (*R*/*S*)-**2** remains (through a degree of microscopic reversibility) comparable to the oxidation of (*R*/*S*)-**3** to *α*-tetralone. (This issue is reminiscent of the "experimental problem" reviewed by Van Gunsteren and colleagues,[74] whereby experimental data is sometimes scarcer than desirable when setting up biomolecular simulations). We believe this assumption to be reasonable in this case, because in each of the two 'chair-flip' conformers of *α*-tetralone, the carbonyl and three aliphatic carbons are likely to have a near-identical arrangement to atoms C1-C4 (Chart 1) in *RS*-**1** or *SR*-**1** (there would be the same such correspondence between *S*-**3** and *SSR*-**2**/*SRS*-**2**; and between *R*-**3** and *RSR*-**2**/*RRS*-**2**). The only major (steric) difference would be the axial atoms in **1** and **2** jutting out above and below the plane of the non-carbonyl ring: as a result, we speculate that the **1**/**2** pair could be more susceptible to steric hindrance at its non-reactive (non-carbonyl) end, possibly increasing the degree of stereoselectivity towards **1** compared to *α*-tetralone/**3**.

Towards an Efficient Approach for Understanding Enzymatic Stereocontrol. The procedure that we employ in this study relies on carefully chosen computational methods to dissect the influence of **FI**-**FIII**—something that is very hard to achieve experimentally. We have deliberately used protocols that (1) require input of the WT structure only; (2) use relatively limited computational resources (short simulations and semiempirical QM treatment); and (3) can be automated. We thus envisage that these protocols can be used in a way that is conceptually similar to the 'high-throughput–multiple independent MD simulations' (HTMI-MD) approach used by Wijma and coworkers[24, 75-76] to efficiently screen newly suggested stereoselective enzyme variants *in silico*, prior to assessment by experiment. We note that the HTMI-MD approach only assesses **FI**, whereas we have shown here that, in particular, **FIII** can be crucial to understand stereoselectivity in certain ketoreductase variants (P94L *act*KR).

The protocols serve to obtain prediction of and insight into stereoselectivity in novel mutants relatively rapidly, and could thus aid the design of novel mutants. For the use of such tools to become more widespread, and to allow screening a large number of enzyme variants, the computational time and resources required should be modest. In this respect, the fact that our relatively low-cost protocols have been able to characterize the stereoselectivity of WT, P94L and V151L *act*KR with encouraging matches to experimental data shows promise for future applications to other enzymes and mutants. In this subsection, we discuss the rationale for the protocols to assess **FI**-**FIII** and evaluate if computational time can be reduced further.

To study **FI** (and to later generate trajectories for **FII** and **FIII**) we employed classical MD simulations, previously employed in several examples of interest with other KRs and SDRs.[10, 77] In contrast to many previous studies that have used a small number of longer MD simulations to collect their data,[10, 26, 77-78] our strategy was to run a large set of independent MD simulations of shorter length (8 × 12 ns, essentially becoming 32 × 12 ns due to the four active sites) to maximize conformational sampling. This length is still significantly greater than those employed by Wijma *et al.* (20-40 × 10 ps or 10 × 100 ps),[24, 76] because this may be necessary to capture relevant enzyme conformational changes. The importance of carrying out multiple independent MD replicas (at least 5-10 as a rule-of-thumb) has been recently restated.[79]

In the case of ketoreductases/SDRs, the α6-α7 loop can adopt alternative conformations (Figure 1c and d), and was proposed to play a role in determining reactivity and selectivity in a different enzyme.[10] We thus confirmed (*vide infra*) that the 12 ns simulation length is sufficient for "closed" chains (B, D) to sample open α6-α7 loop conformations and *vice versa* for "open" chains (A, C). We further explored the possibility of investigating **FI** with simulations shorter than 12 ns, and our data (*Supporting Information*, Figure S4) show that this is feasible: trends are identical with simulations shorter than 0.3 ns, leaving the conclusions unaltered.

The MM/PBSA option was chosen to study **FII** in light of its relatively modest computational cost, as well as its good performance in previous ΔG_{bind}-prediction problems across several enzyme-ligand systems.[32, 80-81] Alternative options such as absolute binding free energy calculations using alchemical perturbations[30-31] would have raised the computational cost significantly.[82] Our decision to run MM/PBSA calculations on a large set of independent MD simulations was to maximize conformational sampling, and has been shown to improve performance.[83-84] In fact, though in our case we employed 12 ns MD replicas to adequately sample conformational changes in the α6-α7 loop, previous work shows that much shorter MD simulations (in the order of hundreds of ps to a few ns) are typically sufficient.[83-84] In our case, tests suggest that using MD trajectories shorter than 4-8 ns may have repercussions on accuracy (*Supporting Information*, Figure S5); likely in part due to the use of fewer snapshots overall. (The 24 ps-interval at which snapshots are extracted is roughly in line with previous MM/PBSA and MM/GBSA indications, but could be reduced).[83-85]

To study **FIII**, we employed a semiempirical QM/MM approach, to avoid specific parameterization of models (*e.g.* as required for empirical valence bond methods)[36] and limit computational cost (*e.g.* as compared to QM/MM or a QM cluster approach with DFT methods). To further limit this cost, we chose to use a small QM region with only groups directly electronically involved in the rate-limiting step (Figure 3), rather than a region that incorporates possible further proton transfers (as used in previous QM/MM studies of *act*KR-related enzymes).[10, 15] A wide range of options is available for calculating reaction barriers with QM/MM.[37-38] In our case, it was not known *a priori* what the orientation of the substrate would be in the different variants (due to its small size in comparison to the relatively large active site in *act*KR). Therefore, an approach using QM/MM optimization alone was not ideal, since both the initial choice of substrate orientation and enzyme conformation may have a large influence on the resulting activation energy barriers, and this would have meant collecting a very large number of starting conformations.[86] We thus opted for QM/MM MD reaction simulations to allow sampling of many orientations, as well as full consideration of entropic effects. Conformational sampling was further enhanced by performing 10 independent reaction simulations (leading to 20

ps sampling per umbrella sampling window), for each of the four possible reactions, across the three *act*KR variants (120 in total); indeed, our tests with 36 independent simulations (3 per case, 6 ps per window; *Supporting Information*, Figure S6) show that results become significantly less reliable (with larger errors).

Notably, QM/MM simulations used a range of starting structures from the many independent classical MD simulations performed: these capture a range of α6-α7 loop conformations (between 'open' and 'closed'). As mentioned earlier, in contrast to previous work on *Lactobacillus kefir* short-chain alcohol dehydrogenase,[10] we found little correlation between the loop openness (as measured by the average of three interatomic loop-enzyme distances) and barrier height (*Supporting Information*, Figures S7 and S8), indicating that in this case, that the α6-α7 loop conformation has little or no influence on *act*KR's catalytic reactivity when turning over small substrates.

4. Conclusions

Biocatalyst (stereo)selectivity can arise due to a combination of effects, including substrate binding (and its affinity in different orientations), as well as reaction barrier. Such effects, and their interplay, can be subtle, especially when dealing with a relatively open active site (as in the case of *act*KR, the enzyme studied here). One route towards stereoselective enzymes (for use in biocatalysis) can thus be to reduce their active site volume (*e.g.* in such a way that substrates can only bind in specific orientations).[24, 78] However, such changes may also reduce their capability for turning over a range of substrates (of different sizes) as well as reduce efficiency.

Here, we have presented a detailed computational study of three factors that may affect stereocontrol in the wild-type ketoreductase *act*KR and in two key variants with similar efficiency (P94L, V151L). We separately investigated the formation (**FI**), the binding free energy (**FII**) and the reactivity (free energy barrier, **FIII**) of reactive complexes, using classical MD simulations, MM/PBSA calculations and QM/MM MD reaction simulations, respectively. We demonstrate that the strict stereoselectivity observed in the P94L and V151L *act*KR variants (both leading to only a small decrease in active site volume) is arising through different mechanisms in each case. For the *S*-selective P94L *act*KR variant, stereoselectivity is driven by the relative difference between activation free energy barriers (**FIII**), whereas for the *R*-selective V151L *act*KR variant, the formation and binding affinity of reactive substrate poses (**FI**, **FII**) are mainly responsible.

The observation that the main effect (or factor) determining stereocontrol can differ from mutant to mutant has important implications for obtaining stereoselective enzymes through redesign: either substrate orientation or efficiency of catalysis (*via* transition state stabilization), or both, may be either altered to confer stereoselectivity.

Our simulations further indicate that each enantiomer of the substrate *trans*-1-decalone (**1**) has a preference for particular reactive poses: *pro*-*S* poses are predominantly found for *SR*-**1**, and *pro*-*R* poses for *RS*-**1**. We note that this was observed in all three *act*KR variants studied, including the mildly *S*-selective WT *act*KR. For the stereoselective reduction of **1**, it is thus the combined effects of the *act*KR variant and substrate enantiomer that determine the final product, with P94L favoring formation of *SSR*-**2** and V151L favoring formation of *RRS*-**2**.

The complexity of our findings is representative of the multifaceted origins of catalysis and selectivity in enzyme biocatalysts, which arises from a subtle interplay of steric, dynamic, and electronic effects. Detailed biomolecular simulation provides valuable means to break down the causes of stereocontrol (*e.g.* in terms of the individual factors **FI**-**FIII**). The computational procedures we have used here to determine the origins of stereocontrol in *act*KR are deliberately generically applicable and employ limited computer resources. They can thus be used for similar evaluations of different enzyme biocatalysts in a time- and cost-effective manner (especially in combination with further automation); although it should be noted that evaluation of **FIII** does require the reaction mechanism to be known in advance. Indeed, we believe that an accurate, yet computationally efficient assessment of the different effects (or factors) dictating an enzyme's stereocontrol is highly valuable for the (re)design of biocatalysts with enhanced stereoselectivity or -specificity, as well as for in-depth understanding of selectivity in existing enzyme variants.

ASSOCIATED CONTENT

Supporting Information. Details on the construction of starting structures, force-field parameters and MD simulations (restraints and equilibration procedure); additional details for QM/MM MD umbrella sampling and QM/MM benchmarking; error analysis; *Gaussian09* `*.log` files for all optimized stationary points obtained from benchmarking (available at https://doi.org/10.19061/iochem-bd-6-12); starting structures and forcefield topologies for the six *act*KR-(NADPH)$_4$-**1**$_4$ systems (`*.zip` folder). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* E-mail: Marc.VanderKamp@bristol.ac.uk

Orcid

Marc W. van der Kamp: 0000-0002-8060-3359
Stefano A. Serapian: 0000-0003-0122-8499

Author Contributions

Both authors designed the experiments, SAS performed experiments, MWK devised the project. Both authors analyzed results, wrote the manuscript and have given approval to the final version of the manuscript.

Funding Sources

MWK is a BBSRC David Phillips Fellow, and the authors wish to thank BBSRC for funding (grant BB/M026280/1).

ACKNOWLEDGEMENT

The authors thank Prof. Shiou-Chan Tsai (Sheryl) and Gabriel O. Moreno (*UC Irvine*), as well as Prof. Matt Crump (*Bristol*), for helpful discussions. This work was conducted using the computational facilities of the Advanced Computing Research Centre of the University of Bristol.

REFERENCES

1. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A., Type II Polyketide Synthases: Gaining a Deeper Insight into Enzymatic Teamwork. *Nat. Prod. Rep.* **2007,** *24*, 162-190.
2. Zhang, Z.; Pan, H. X.; Tang, G. L., New Insights into Bacterial Type II Polyketide Biosynthesis. *F1000Research* **2017,** *6*, 172.
3. Du, D.; Katsuyama, Y.; Shin-ya, K.; Ohnishi, Y., Reconstitution of a Type II Polyketide Synthase That Catalyzes Polyene Formation. *Angew. Chem. Int. Ed.* **2018,** *57*, 1954-1957.
4. Javidpour, P.; Das, A.; Khosla, C.; Tsai, S.-C., Structural and Biochemical Studies of the Hedamycin Type II Polyketide Ketoreductase (HedKR): Molecular Basis of Stereo- and Regiospecificities. *Biochem.* **2011,** *50*, 7426-7439.
5. Korman, T. P.; Hill, J. A.; Vu, T. N.; Tsai, S.-C., Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochem.* **2004,** *43*, 14529-14538.
6. Tang, Y.; Lee, H. Y.; Tang, Y.; Kim, C.-Y.; Mathews, I.; Khosla, C., Structural and Functional Studies on SCO1815: A B-Ketoacyl-Acyl Carrier Protein Reductase from Streptomyces Coelicolor A3(2). *Biochem.* **2006,** *45*, 14085-14093.
7. Valentic, T. R.; Jackson, D. R.; Brady, S. F.; Tsai, S.-C., Comprehensive Analysis of a Novel Ketoreductase for Pentangular Polyphenol Biosynthesis. *ACS Chem. Biol.* **2016,** *11*, 3421-3430.
8. Moore, J. C.; Pollard, D. J.; Kosjek, B.; Devine, P. N., Advances in the Enzymatic Reduction of Ketones. *Acc. Chem. Res.* **2007,** *40*, 1412-1419.
9. Huisman, G. W.; Liang, J.; Krebber, A., Practical Chiral Alcohol Manufacture Using Ketoreductases. *Curr. Opin. Chem. Biol.* **2010,** *14*, 122-129.
10. Noey, E. L.; Tibrewal, N.; Jiménez-Osés, G.; Osuna, S.; Park, J.; Bond, C. M.; Cascio, D.; Liang, J.; Zhang, X.; Huisman, G. W.; Tang, Y.; Houk, K. N., Origins of Stereoselectivity in Evolved Ketoreductases. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112*, E7065-E7072.
11. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the Third Wave of Biocatalysis. *Nature* **2012,** *485*, 185.
12. Bornscheuer, U. T., The Fourth Wave of Biocatalysis Is Approaching. *Phil. Trans. R. Soc. A* **2018,** *376*, 20170063.
13. Hadfield, A. T.; Limpkin, C.; Teartasin, W.; Simpson, T. J.; Crosby, J.; Crump, M. P., The Crystal Structure of the ActIII Actinorhodin Polyketide Reductase: Proposed Mechanism for ACP and Polyketide Binding. *Struct.* **2004,** *12*, 1865-1875.
14. Korman, T. P.; Tan, Y.-H.; Wong, J.; Luo, R.; Tsai, S.-C., Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase. *Biochem.* **2008,** *47*, 1837-1847.
15. Medina, F. E.; Neves, R. P. P.; Ramos, M. J.; Fernandes, P. A., A QM/MM Study of the Reaction Mechanism of Human [Small Beta]-Ketoacyl Reductase. *Phys. Chem. Chem. Phys.* **2017,** *19*, 347-355.
16. Javidpour, P.; Korman, T. P.; Shakya, G.; Tsai, S.-C., Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase. *Biochem.* **2011,** *50*, 4638-4649.
17. Javidpour, P.; Bruegger, J.; Srithahan, S.; Korman, Tyler P.; Crump, Matthew P.; Crosby, J.; Burkart, Michael D.; Tsai, S.-C., The Determinants of Activity and Specificity in Actinorhodin Type II Polyketide Ketoreductase. *Chem. Biol.* **2013,** *20*, 1225-1234.
18. Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; Van der Kamp, M. W., Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018,** *0*, e1393-e1415.
19. Daniels, A. D.; Campeotto, I.; Van der Kamp, M. W.; Bolt, A. H.; Trinh, C. H.; Phillips, S. E. V.; Pearson, A. R.; Nelson, A.; Mulholland, A. J.; Berry, A., Reaction Mechanism of N-Acetylneuraminic Acid Lyase Revealed by a Combination of Crystallography, QM/MM Simulation, and Mutagenesis. *ACS Chem. Biol.* **2014,** *9*, 1025-1032.
20. Kulkarni, Y. S.; Liao, Q.; Petrović, D.; Krüger, D. M.; Strodel, B.; Amyes, T. L.; Richard, J. P.; Kamerlin, S. C. L., Enzyme Architecture: Modeling the Operation of a Hydrophobic Clamp in Catalysis by Triosephosphate Isomerase. *J. Am. Chem. Soc.* **2017,** *139*, 10514-10525.
21. Raich, L.; Borodkin, V.; Fang, W.; Castro-López, J.; Van Aalten, D. M. F.; Hurtado-Guerrero, R.; Rovira, C., A Trapped Covalent Intermediate of a Glycoside Hydrolase on the Pathway to Transglycosylation. Insights from Experiments and Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.* **2016,** *138*, 3325-3332.
22. Amrein, B. A.; Steffen-Munsberg, F.; Szeler, I.; Purg, M.; Kulkarni, Y.; Kamerlin, S. C. L., CADEE: Computer-Aided Directed Evolution of Enzymes. *IUCrJ* **2017,** *4*, 50-64.
23. Hediger, M. R.; De Vico, L.; Rannes, J. B.; Jäckel, C.; Besenmatter, W.; Svendsen, A.; Jensen, J. H., In Silico Screening of 393 Mutants Facilitates Enzyme Engineering of Amidase Activity in CalB. *PeerJ* **2013,** *1*, e145.
24. Wijma, H. J.; Floor, R. J.; Bjelic, S.; Marrink, S. J.; Baker, D.; Janssen, D. B., Enantioselective Enzymes by Computational Design and in Silico Screening. *Angew. Chem.* **2015,** *127*, 3797-3801.
25. Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S., Computational Tools for the Evaluation of Laboratory-Engineered Biocatalysts. *Chem. Commun.* **2017,** *53*, 284-297.
26. Rinaldi, S.; Van der Kamp, M. W.; Ranaghan, K. E.; Mulholland, A. J.; Colombo, G., Understanding Complex Mechanisms of Enzyme Reactivity: The Case of Limonene-1,2-Epoxide Hydrolases. *ACS Catal.* **2018,** *8*, 5698-5707.
27. By "reactive pose", we mean an enzyme-substrate complex positioned such that the substrate is in place for the chemical reaction to start; others have recently referred to this as a "near attack conformation" or NAC. See *e.g.* ref. 24.
28. Janfalk Carlsson, Å.; Bauer, P.; Dobritzsch, D.; Nilsson, M.; Kamerlin, S. C. L.; Widersten, M., Laboratory-Evolved Enzymes Provide Snapshots of the Development of Enantioconvergence in Enzyme-Catalyzed Epoxide Hydrolysis. *ChemBioChem* **2016,** *17*, 1693-1697.
29. Sun, Z.; Lonsdale, R.; Kong, X.-D.; Xu, J.-H.; Zhou, J.; Reetz, M. T., Reshaping an Enzyme Binding Pocket for Enhanced and Inverted Stereoselectivity: Use of Smallest Amino Acid Alphabets in Directed Evolution. *Angew. Chem. Int. Ed.* **2015,** *54*, 12410-12415.
30. Woods, C., J.; Malaisree, M.; Hannongbua, S.; Mulholland, A. J., A Water-Swap Reaction Coordinate for the Calculation of Absolute Protein–Ligand Binding Free Energies. *J. Chem. Phys.* **2011,** *134*, 054114.
31. Cabeza de Vaca, I.; Qian, Y.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L., Enhanced Monte Carlo Methods for Modeling Proteins Including Computation of Absolute Free Energies of Binding. *J. Chem. Theory Comput.* **2018,** *14*, 3279-3288.
32. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015,** *10*, 449-461.
33. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33*, 889-897.
34. Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M., Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010,** *31*, 797-810.
35. Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T.-B. N.; Wang, H.; Ren, P.; Luo, R., Calculating Protein–Ligand Binding Affinities with MMPBSA: Method and Error Analysis. *J. Comput. Chem.* **2016,** *37*, 2436-2446.

36. Kamerlin, S. C. L.; Warshel, A., The Empirical Valence Bond Model: Theory and Applications. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011,** *1*, 30-45.
37. Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R., Multiscale Modeling of Enzymes: QM-Cluster, QM/MM, and QM/MM/MD: A Tutorial Review. *Int. J. Quant. Chem.* **2018,** *118*, e25558-e25591.
38. Van der Kamp, M. W.; Mulholland, A. J., Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochem.* **2013,** *52*, 2708-2728.
39. Warshel, A.; Levitt, M., Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976,** *103*, 227-249.
40. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011,** *7*, 2284-2295.
41. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.
42. Case, D. A.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Götz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; Le Grand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A. E.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J. M.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber 2017*, University of California, San Francisco: 2017.
43. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with Amber on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9*, 3878-3888.
44. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.
45. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996,** *14*, 33-38.
46. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
47. Holmberg, N.; Ryde, U.; Bulow, L., Redesign of the Coenzyme Specificity in L-Lactate Dehydrogenase from Bacillus Stearothermophilus Using Site-Directed Mutagenesis and Media Engineering. *Protein Eng.* **1999,** *12*, 851-856.
48. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.
49. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004,** *25*, 1157-1174.
50. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000,** *21*, 132-146.
51. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002,** *23*, 1623-1641.
52. Crowley, M.; Darden, T.; Cheatham, T.; Deerfield, D., Adventures in Improving the Scaling and Accuracy of a Parallel Molecular Dynamics Program. *J. Supercomput.* **1997,** *11*, 255-278.
53. Darden, T.; York, D. M.; Pedersen, L. G., Particle Mesh Ewald: An N·Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
54. Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.
55. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N′-Methylamide. *Biopolymers* **1992,** *32*, 523-535.
56. Shuichi, M.; A., K. P., Settle: An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992,** *13*, 952-962.
57. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012,** *8*, 3314-3321.
58. Tan, C.; Yang, L.; Luo, R., How Well Does Poisson−Boltzmann Implicit Solvent Agree with Explicit Solvent? A Quantitative Analysis. *J. Phys. Chem. B* **2006,** *110*, 18680-18687.
59. Duan, L.; Liu, X.; Zhang, J. Z. H., Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein–Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016,** *138*, 5722-5728.
60. Walker, R. C.; Crowley, M. F.; Case, D. A., The Implementation of a Fast and Accurate QM/MM Potential Method in Amber. *J. Comput. Chem.* **2008,** *29*, 1019-1031.
61. Seabra, G. d. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E., Implementation of the SCC-DFTB Method for Hybrid QM/MM Simulations within the Amber Molecular Dynamics Package. *J. Phys. Chem. A* **2007,** *111*, 5655-5664.
62. Stewart, J. J. P., Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007,** *13*, 1173-1213.
63. Grimme, S., Improved Second-Order Møller–Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003,** *118*, 9095-9102.
64. Grimme, S.; Goerigk, L.; Fink, R. F., Spin-Component-Scaled Electron Correlation Methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2*, 886-906.
65. Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.
66. A collection of relevant calculations is available on the ioChemBD server. https://doi.org/10.19061/iochem-bd-6-12. Last accessed: December 3rd 2018.
67. Álvarez-Moreno, M.; de Graaf, C.; López, N.; Maseras, F.; Poblet, J. M.; Bo, C., Managing the Computational Chemistry Big Data Problem: The ioChem-BD Platform. *J. Chem. Inf. Model.* **2015,** *55*, 95-103.
68. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992,** *13*, 1011-1021.
69. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995,** *16*, 1339-1350.
70. Grossfield, A. *WHAM: The Weighted Histogram Analysis Method*, version 2.0.9.1; 2013.
71. Di Maio, G.; Gabriella Mascia, M.; Vecchi, E., Does Substituent's Conformation Influence the Kinetics of Reduction Reactions on Trans-4-X-Decal-1-Ones and to What Extent? *Tetrahedron* **2002,** *58*, 3313-3318.
72. Wu, Y. D.; Tucker, J. A.; Houk, K. N., Stereoselectivities of Nucleophilic Additions to Cyclohexanones Substituted by Polar Groups. Experimental Investigation of Reductions of Trans-Decalones and Theoretical Studies of Cyclohexanone Reductions. The Influence of Remote Electrostatic Effects. *J. Am. Chem. Soc.* **1991,** *113*, 5018-5027.
73. Østergaard, L. H.; Kellenberger, L.; Cortés, J.; Roddis, M. P.; Deacon, M.; Staunton, J.; Leadlay, P. F., Stereochemistry of Catalysis by the Ketoreductase Activity in the First Extension Module of the Erythromycin Polyketide Synthase. *Biochem.* **2002,** *41*, 2719-2726.
74. Van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.;

Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B., Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chem. Int. Ed.* **2006,** *45*, 4064-4092.

75. Li, R.; Wijma, H. J.; Song, L.; Cui, Y.; Otzen, M.; Tian, Y. e.; Du, J.; Li, T.; Niu, D.; Chen, Y.; Feng, J.; Han, J.; Chen, H.; Tao, Y.; Janssen, D. B.; Wu, B., Computational Redesign of Enzymes for Regio- and Enantioselective Hydroamination. *Nat. Chem. Biol.* **2018,** *14*, 664-670.

76. Wijma, H. J.; Marrink, S. J.; Janssen, D. B., Computationally Efficient and Accurate Enantioselectivity Modeling by Clusters of Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2014,** *54*, 2079-2092.

77. Mugnai, M. L.; Shi, Y.; Keatinge-Clay, A. T.; Elber, R., Molecular Dynamics Studies of Modular Polyketide Synthase Ketoreductase Stereospecificity. *Biochem.* **2015,** *54*, 2346-2359.

78. Sun, Z.; Wu, L.; Bocola, M.; Chan, H. C. S.; Lonsdale, R.; Kong, X.-D.; Yuan, S.; Zhou, J.; Reetz, M. T., Structural and Computational Insight into the Catalytic Mechanism of Limonene Epoxide Hydrolase Mutants in Stereoselective Transformations. *J. Am. Chem. Soc.* **2018,** *140*, 310-318.

79. Knapp, B.; Ospina, L.; Deane, C. M., Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018,** *14*, 6127–6138.

80. Kuhn, B.; Kollman, P. A., Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000,** *43*, 3786-3791.

81. Wang, J.; Morin, P.; Wang, W.; Kollman, P. A., Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of Tibo Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001,** *123*, 5221-5230.

82. Aldeghi, M.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Statistical Analysis on the Performance of Molecular Mechanics Poisson–Boltzmann Surface Area Versus Absolute Binding Free Energy Calculations: Bromodomains as a Case Study. *J. Chem. Inf. Model.* **2017,** *57*, 2203-2221.

83. Adler, M.; Beroza, P., Improved Ligand Binding Energies Derived from Molecular Dynamics: Replicate Sampling Enhances the Search of Conformational Space. *J. Chem. Inf. Model.* **2013,** *53*, 2065-2072.

84. Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V., Accurate Ensemble Molecular Dynamics Binding Free Energy Ranking of Multidrug-Resistant HIV-1 Proteases. *J. Chem. Inf. Model.* **2010,** *50*, 890-905.

85. Genheden, S.; Ryde, U., How to Obtain Statistically Converged MM/GBSA Results. *J. Comput. Chem.* **2010,** *31*, 837-846.

86. Ryde, U., How Many Conformations Need to Be Sampled to Obtain Converged QM/MM Energies? The Curse of Exponential Averaging. *J. Chem. Theory Comput.* **2017,** *13*, 5745-5752.